# Tableau with Cloudera Hadoop Connector –

## Prerequisites –

1. Install Oracle VirtualBox or VMware player.
2. Download Cloudera quickstart VM from –<http://www.cloudera.com/content/cloudera/en/downloads/quickstart_vms/cdh-5-4-x.html>
3. Download the hive ODBC connector drivers from cloudera website (select the version according to the host Windows machine where Tableau will be running) -  
   <http://www.cloudera.com/content/cloudera/en/downloads/connectors/hive/odbc/hive-odbc-v2-5-16.html>
4. Download the latest Tableau Desktop trial version (this works for windows OS only) from -

<http://www.tableau.com/products/trial>

## Installation –

Getting started with a Cloudera VM on VirtualBox, follow the installation instructions from DataCafe Git page –

<https://github.com/eenov8/Apache-Hadoop/blob/master/Getting%20started%20with%20CDH.docx?raw=true>

Installation of ODBC Hive connector is simple wizard based only.

Installation of Tableau needs to be completed with default selections only.

Connecting Tableau with Cloudera Hadoop VM using hiveserver2 –

## Load data in the Hive table in CDH –

1. Download the data from –

<https://github.com/eenov8/Apache-Hadoop/blob/master/superstore.txt>

1. Copy the file to HDFS –

[cloudera@quickstart ~]$ hadoop fs –copyFromLocal Downloads/superstore.txt /tableau/

1. Using the hue interface load the data in hive.

Get the detailed steps from this video on YouTube –

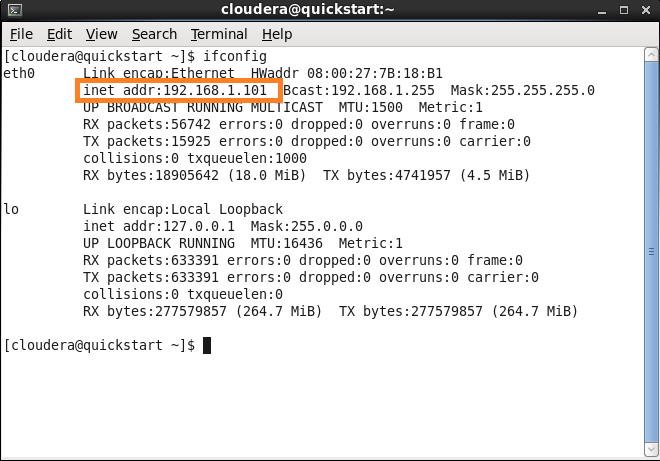
<https://youtu.be/RFm0CH0Zdnw>

1. Start hiveserver2 on cloudera Hadoop VM –

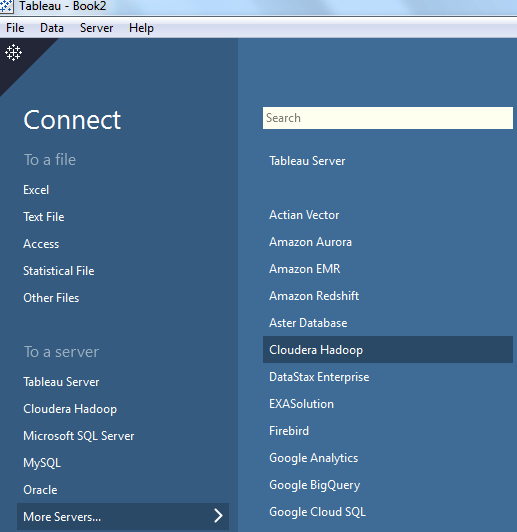
[cloudera@quickstart ~]$ sudo service hive-server2 start

## Create Tableau connection with our cloudera VM

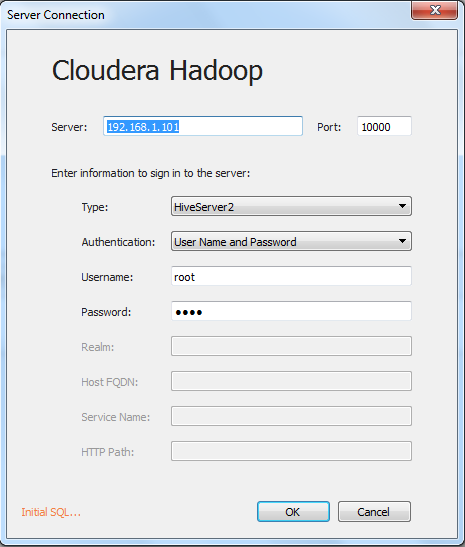
1. Get the IP address of cloudera Hadoop VM



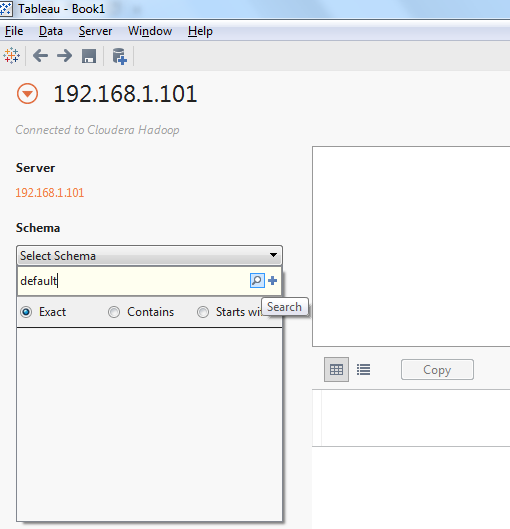
1. Create a new Tableau project and connect to cloudera hadoop as below –



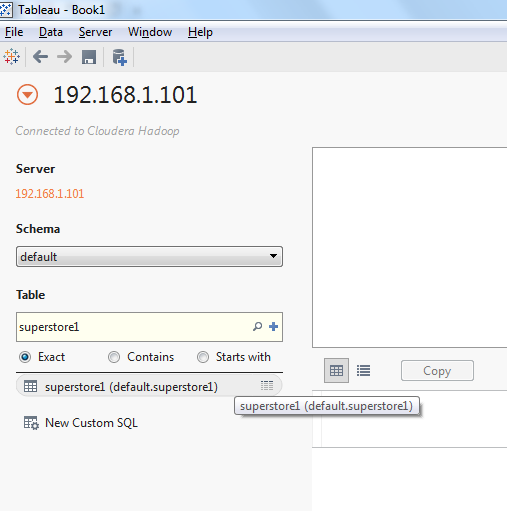
1. Fill in the details as (root/root) –



1. Select the default schema to which we have created the table above -



1. Select the table –



1. Now go to Sheet1 and create your own visualizations.